

La proyección de los repositorios institucionales

3er Encuentro OS-Repositorios. Facultad de Informática, Universidad Complutense
Madrid 10-12 Diciembre 2008

Estadísticas de uso de repositorios institucionales: ¿un indicador fiable de comparación?

Pablo de Castro Martín
Oficina Técnica Digital.CSIC
digital.csic@bib.csic.es

Agradecimientos:
Juan Corrales (Consortio Madroño)
Julio Pemau (Digital.CSIC)

Indicadores del avance de la OAI: ¿sólo cantidad?

OAIster
...find the pearls

Home Search Help • About Using OAIster News

About OAIster

OAIster currently provides access to **18,362,804** records from **1035** contributors.

OAIster is a **union catalog of digital resources**. We provide access to these digital resources by "harvesting" their descriptive metadata (records) using [OAI-PMH](#) (the Open Archives Initiative Protocol for Metadata Harvesting).

[More...](#)

[View Data Contributors](#)

[Collection Development Policy](#)

Using OAIster

OAIster can be searched by Title, Author/Creator, Subject, Language or Entire Record. Searches can also be limited by resource type (text, image, audio, video, dataset) and sorted by title, author, date and hit frequency. Results allow further limiting by data contributor (i.e., where the record was harvested from).

[More...](#)

SEARCH

[Search Help](#)

[View Data Contributors](#)

News & Updates

2008 Sep 5
Enjoy a new, cleaner version of the OAIster data. Many new repositories have been added, and many dead links were removed during August.

2008 Jul 17
Read our recent article: [Google Still Not Indexing Hidden Web URLs](#).

[More...](#)

Last update: 31 Oct 2008
Next update: 5 Nov 2008

New data contributors harvested:

Estadísticas actualmente proporcionadas por agregadores de repositorios

RECOLECTA

Recolector de ciencia abierta

[Inicio](#)

Buscador

- [Búsqueda simple](#)
- [Búsqueda avanzada](#)
- [Navegar por recursos](#)
- [Recursos recolectados](#)
- [Administración](#)

[Sherpa/Romeo](#)

[Documentación](#)

[Contacto](#)

Buscador

Obtener todos los registros

Registros de metadatos agrupados por Recurso: **Todos**

Ordenados por: **Fecha**

Página: **[1]** [2] [3] [4] [5] [6] [7] [8] [9] [10] [11] [12] [13] [14] [15] [16] [17] [18] [19] [20]
[Sigs. >>]

Patterns of juvenile habitat use by the spider crab *Maja brachydactyla* as revealed by stable isotope analyses | Patrones de uso del hábitat de la centolla *Maja brachyd...*

Juan Freire; Departamento de Biología Animal, Biología Vegetal e Ecología, Universidade da Coruña | Sirka Carabel; Depart...

Patterns of habitat use by spider crab juveniles, *Maja brachydactyla*, from two geographic areas on the NW coast of the Iberian Peninsula were studied through the analysis of carbon and nitrogen stable

2009-03-30
Scientia Marina
[Registro completo y enlace al documento](#)

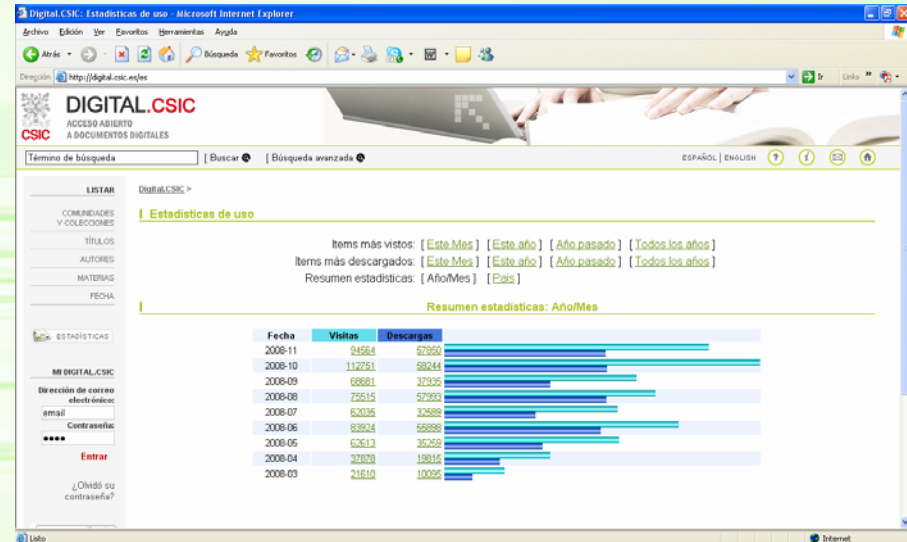
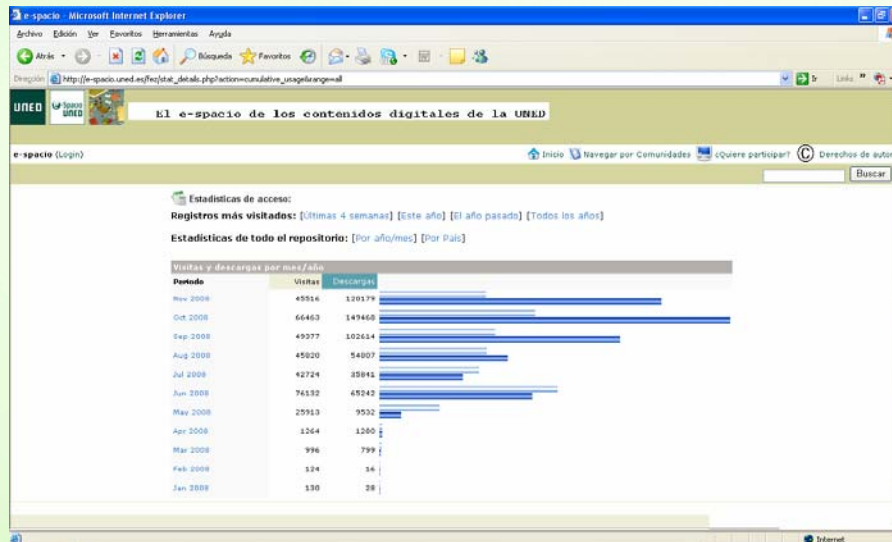
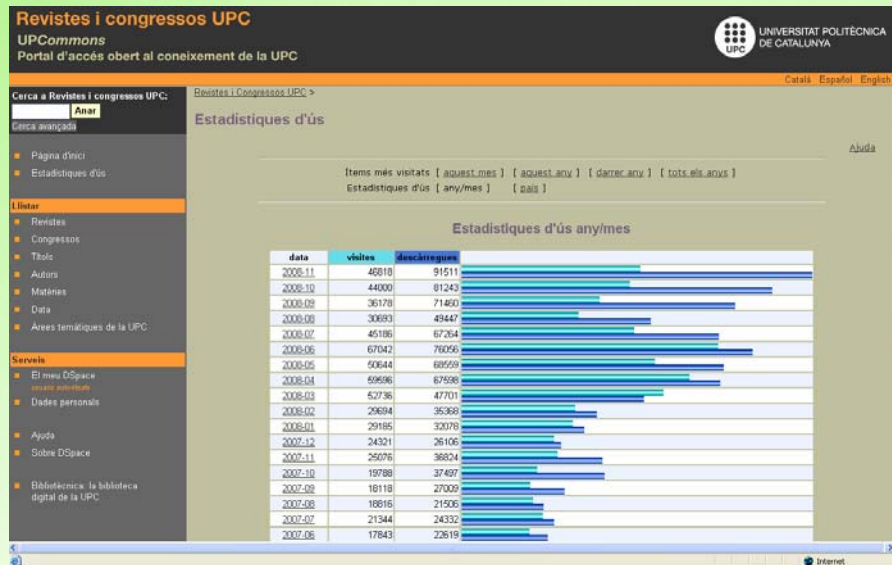
Ayuda

Recursos

Todos (391594)

- Anales del Jardín Botánico (264)
- Anuario de Estudios Americanos (107)
- Anuario de Estudios Medievales (79)
- Arbor (342)
- Archivo Digital UPM (990)
- Archivo Español de Arqueología (38)
- Archivo Institucional E-Prints C... (5419)
- Archivo digital del Departamento... (5)
- Archivos de Medicina (57)
- Asclepio (320)
- BURJC-DIGITAL Universidad Rey Ju... (192)

Progresiva implantación de módulos de estadísticas



¿Podrían las estadísticas de uso proponerse como indicadores de la calidad de los repositorios institucionales?



Proposal to webmasters for 2009 edition

Usage data could be very useful for the Ranking Web of World Repositories as it has been suggested by most of the academic community. It is difficult to obtain data from every repository and in fact most of the figures are not comparable because there is no standard yet about statistics collection. We suggest using a common, simple and free system to solve, at least partially, this problem. We ask to webmasters to install Google Analytics (<http://www.google.com/analytics>) at their websites. We will then be able to kindly ask you for the statistics collected during the period from July 1st -December 31st so this data can be used for the January 2009 edition of the Ranking.


Problemas en la definición de estándares

1. Heterogeneidad de plataformas de trabajo

REPOSITORIO	SOFTWARE	MÓDULO ESTADÍSTICAS
UPCommons: E-Prints UPC	DSpace	Módulo "Tasmania"
UPCommons: Revistes i Congresos	Dspace	Módulo "Tasmania"
RUA (Universidad de Alicante)	DSpace	Módulo "Tasmania"
e-Spacio UNED	Fedora	
e-archivo UC3M	DSpace	Módulo "Minho"
Digital.CSIC	DSpace	Módulo "Tasmania"
Archivo Digital UPM	e-Prints	

Problemas en la definición de estándares

2. Filtrado de los datos brutos de acuerdo con criterios estándar

Buscar 

Búsqueda avanzada

¿Qué es?

Estadísticas

Novedades

Lista por

Comunidades y colecciones

Títulos

Autores

Materias


Fecha


Autentica por

Recibir novedades

Mi repositorio

Editar perfil



 Dipòsit de la Recerca de Catalunya



Estadísticas

- [Documentos incorporados por institución](#)
- [Documentos incorporados y recolectados por institución](#)
- [Documentos consultados por institución](#)
- [Documentos consultados por país e institución](#)
- [Los 30 documentos más consultados este año](#)
- [Evolución de RECERCAT](#)



Notas

- Una consulta a un documento se define como un acceso a su página de metadatos, o como la descarga directa de este documento cuando ésta no es a través de su propia página de metadatos.
- Los accesos automáticos de robots, tanto de aquellos que consultan el fichero *robots.txt* como otros detectados por nosotros, no se contabilizan como consultas. También se excluyen los accesos fraudulentos (muchos accesos desde una misma IP) y los accesos realizados desde direcciones del CESA y del CBUC.
- Para determinar las consultas por país se usan las librerías GeoIP.


Todos los documentos incluidos en este repositorio están bajo una licencia Creative Commons


Coord

Patro



Basad



Problemas en la definición de estándares

3. Discrepancias en los indicadores evaluados



Mes	Visitantes distintos	Número de visitas	Páginas	Solicitudes	Tráfico
Ene 2008	13517	21037	147047	265818	150.13 GB
Feb 2008	16687	22875	132299		
Mar 2008	17450	21528	92360		
Abr 2008	23092	27787	104505		
May 2008	31303	36837	103879		
Jun 2008	39262	46629	132264		

Downloads and views - Time Series

Number of downloads and views in the period.

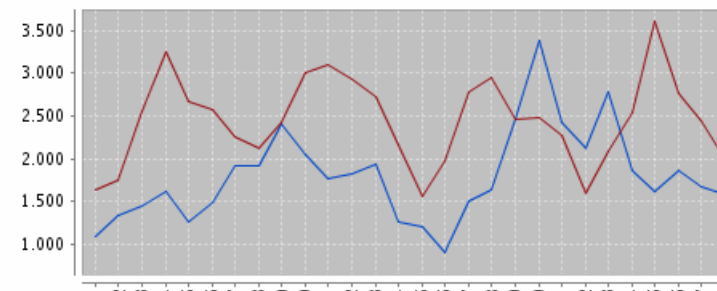
Period: Total :: Year :: Month :: Day

Month/Year

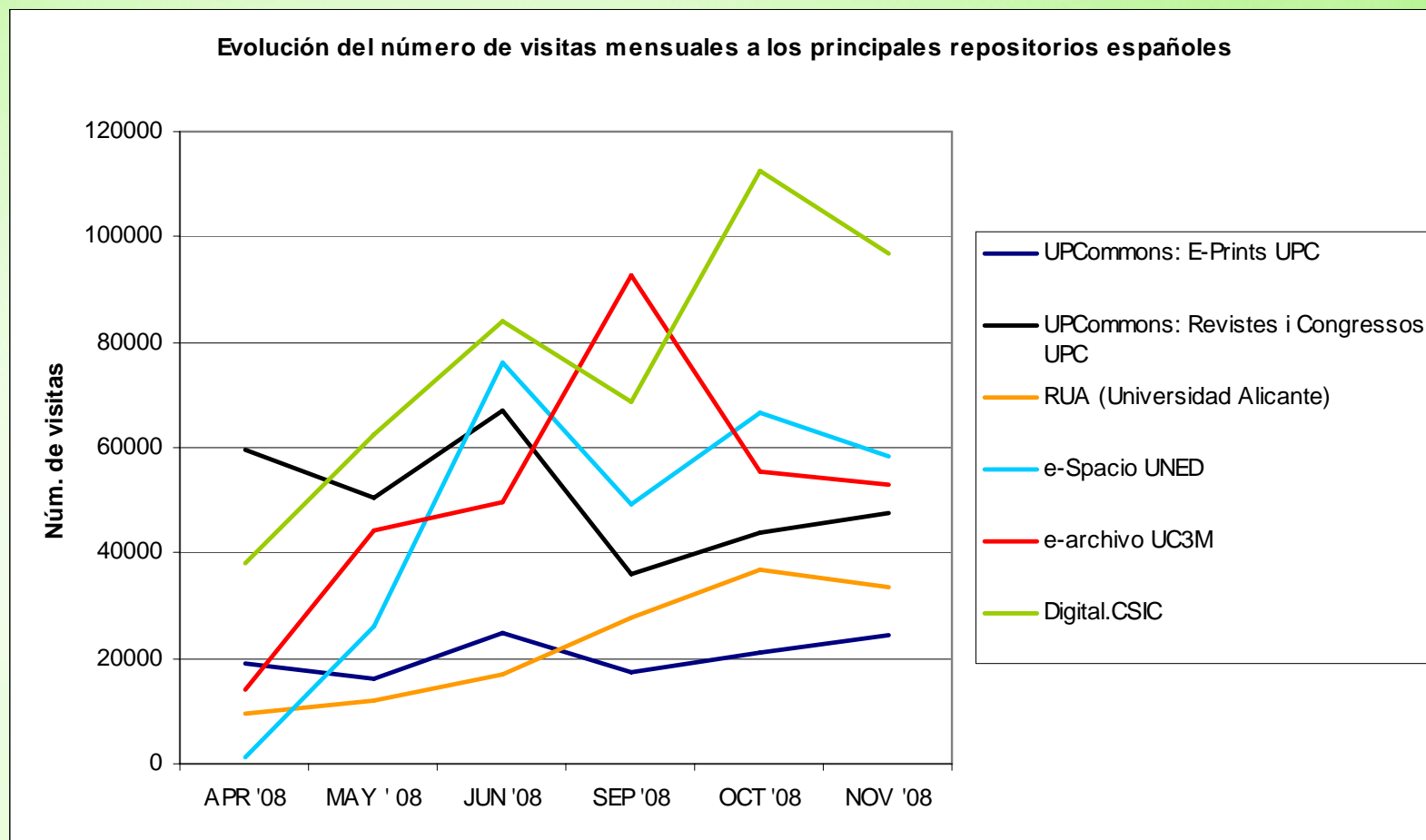
Downloads and views

Day	Downloads	Views
01, sáb	1.636,0	1.083
02, dom	1.750,1	1.327
03, lun	2.562,1	1.450
04, mar	3.253,0	1.620
05, mié	2.663,4	1.255
06, jue	2.562,7	1.481
07, vie	2.247,3	1.913
08, sáb	2.117,2	1.909
09, dom	2.424,9	2.405
10, lun	2.998,1	2.042

Downloads and views per day



Consecuencia: en el momento presente, las estadísticas de uso tienen valor básicamente como indicadores *internos* de evaluación



Sin embargo: Estándares de calidad de estadísticas de uso



Counting Online Usage of NeTworked Electronic Resources

[HOME](#) | [ABOUT COUNTER](#) | [FAQs](#) | [CODE OF PRACTICE](#) | [NEWS AND ACTIVITIES](#) | [COMPLIANT VENDORS](#) | [CONTACT Us](#) | [MEMBERS](#)

About COUNTER

Launched in March 2002, COUNTER (Counting Online Usage of NeTworked Electronic Resources) is an international initiative serving librarians, publishers and intermediaries by setting standards that facilitate the [recording and reporting of online usage statistics in a consistent, credible and compatible way](#). The first COUNTER Code of Practice, covering online journals and databases, was published in 2003. COUNTER's coverage was extended further with the launch of the Code of Practice for online books and reference works in 2006. The body of COUNTER compliant usage statistics has steadily grown as more and more vendors have adopted the COUNTER Codes of Practice. This has contributed to the new discipline of usage bibliometrics and a great deal of work is underway to try to establish value metrics associated with usage, in which the COUNTER compliant statistics play an increasingly important role.

COUNTER does more than just set the standards for usage reports; we are co-operating with a number of organizations to develop a range of usage-related research and services. In 2006 COUNTER carried out research, sponsored by JISC (the UK Joint Information Systems Committee) on the effects of publisher platforms on usage and we are currently collaborating with the UK Serials Group on the possible development of a new Journal Usage Factor metric. Summary reports on both these projects can be found on the COUNTER website at <http://www.projectcounter.org/news.html>. COUNTER has also worked with NISO on SUSHI (Standardised Usage Harvesting Initiative) to develop a protocol to facilitate the automated harvesting and consolidation of usage statistics from different vendors. This protocol may be found on the NISO website at <http://www.niso.org/schemas/sushi/index.html#COUNTER>

Proyectos internacionales sobre estadísticas de uso de repositorios institucionales



- home
- about us
- what we do
- funding opportunities

What we do

- Programmes
- Projects**
- Services
- Themes
- Campaigns

on this page:

- [documents & multimedia](#)
- [email this page](#)

Future objectives

While COUNTER has greatly improved the reliability and usability of online vendor usage statistics, there is still much for us to do, not only to help vendors further improve their usage reports and to help librarians to make sense of them, but also to keep the COUNTER codes up to date with changes in the online delivery of information. Our future objectives fall into three broad categories. First, to improve further the reliability of the core COUNTER data and **extend scope of the Code of Practice beyond journals databases and books**. Second, to continue to increase the number of COUNTER compliant vendors. Third, to work with other industry organizations to facilitate the implementation of COUNTER and develop metrics based on the COUNTER data that are of practical value to both librarians and vendors.

Publisher and Institutional Repository Usage Statistics (PIRUS)

Start date: 1 August 2008
End date: 1 December 2008
Funding programme: [PALS Metadata and Interoperability programme \(phase 3\)](#)
JISC theme(s): [Information environment](#)
Committees: [JISC Integrated Information Environment committee](#)

The aim of this project is to develop COUNTER-compliant usage reports at the individual article level that can be implemented by any entity (publisher, aggregator, IR, etc.,) that hosts online journal articles and will enable the usage of research outputs to be recorded, reported and consolidated at a global level in a standard way.

project staff

Project Manager

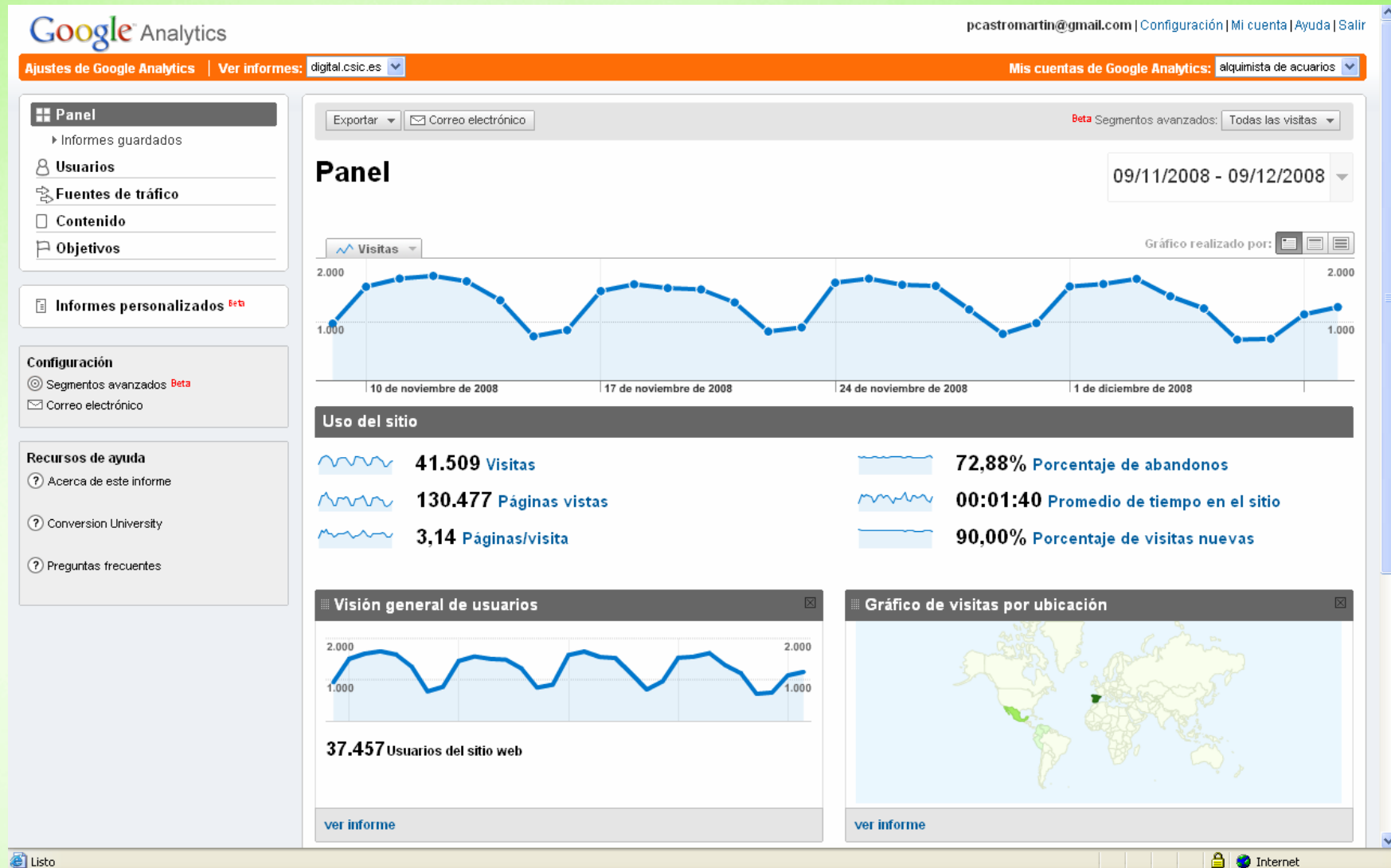
- Peter Shepherd, Counter Project

related projects

- [Blooming CQF-QMT \(Credit Qualification Framework – Question Meta Tags\)](#)
- [OPLE \(ONIX-PL Editor\) Version 1.0](#)

documents and multimedia

¿Y a nivel nacional?



Proyecto del Consorcio Madroño

**10 años**
con la ciencia
y la tecnología

BUSCADOR

CONTACTO

SUSCRIPCIÓN

MAPA

REGISTRO

AYUDA



MADRI+D

INFORMACIÓN I+D

EMPRESAS

CIENCIA Y SOCIEDAD

sistema madri+d

e-ciencia buscador de archivos abiertos

EL PROYECTO

BUSCAR DOCUMENTOS

ARCHIVAR DOCUMENTOS

DERECHOS DE AUTOR

DOCUMENTACIÓN

e-ciencia buscador de archivos abiertos

Transmitir conocimiento a la sociedad



Presentación

La Comunidad de Madrid, en un claro apoyo al movimiento de "acceso abierto" (open access) de la información y producción científica por medios electrónicos, ofrece a la sociedad esta Plataforma digital de acceso libre a la producción científica, generada por las Universidades públicas radicadas en la región madrileña – integradas en el Consorcio Madroño – así como por cualquier investigador de otra universidad, institución, OPI, o de manera independiente, que pretenda y desee participar en la difusión de la ciencia en abierto, es decir, descubrir y transmitir conocimiento a la comunidad científica y a la sociedad en general.

El proyecto E-ciencia, pretende, por tanto:

- Dar soporte a las iniciativas de comunicación científica en acceso abierto a los investigadores interesados, a través de una red de archivos digitales, mediante una búsqueda centralizada.
- Ofrecer un espacio de acceso abierto a las publicaciones científicas generadas por la comunidad científica de la Comunidad de Madrid.
- Incrementar la visibilidad, difusión e impacto de la literatura científica en la Comunidad de Madrid.
- Crear un espacio de difusión, promoción y apoyo al movimiento por el acceso libre a la información científica.
- Ofrecer asesoramiento y guía en las cuestiones relacionadas con la comunicación científica en acceso abierto.
- Investigar, analizar y evaluar el impacto de los cambios en las formas de publicación científica.

Si tiene dudas o sugerencias puede ponerse en contacto con e-ciencia@madrimasd.org.

Instituciones participantes

- Consejería de Educación, madri+d
- Consorcio Madroño
- CSIC
- Biblioteca UAH
- Biblioteca UAM
- Biblioteca UC3M
- Biblioteca UCM
- Biblioteca UNED
- Biblioteca UPM
- Biblioteca URJC

Aplicación para la obtención de estadísticas de uso de repositorios

La herramienta de logs está pensada para sacar las estadísticas de uso de repositorios accedidos a través de un servidor apache y alojados en máquinas linux.

Está pensada para usar sólo software común a prácticamente todas las instalaciones linux. Sólo necesita tener instalado el shell bash y las herramientas egrep y sed.

Para los accesos y las descargas se han tenido en cuenta las recomendaciones de counter de eliminar dobles clicks, crawlers y descargas incompletas.

Es necesario tener el apache configurado con el log combined para conocer los agentes de usuario.

En esta versión extrae estadísticas en ficheros csv por año, mes y día de:

- Número de accesos al repositorio
- Número de descargas de documentos
- Número de IPs únicas desde las que se ha accedido al repositorio
- Número de IPs únicas desde las que se han descargado documentos

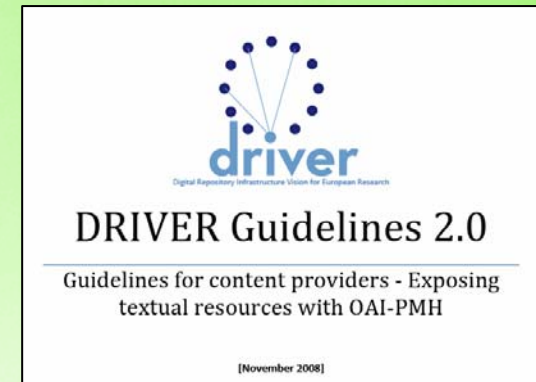
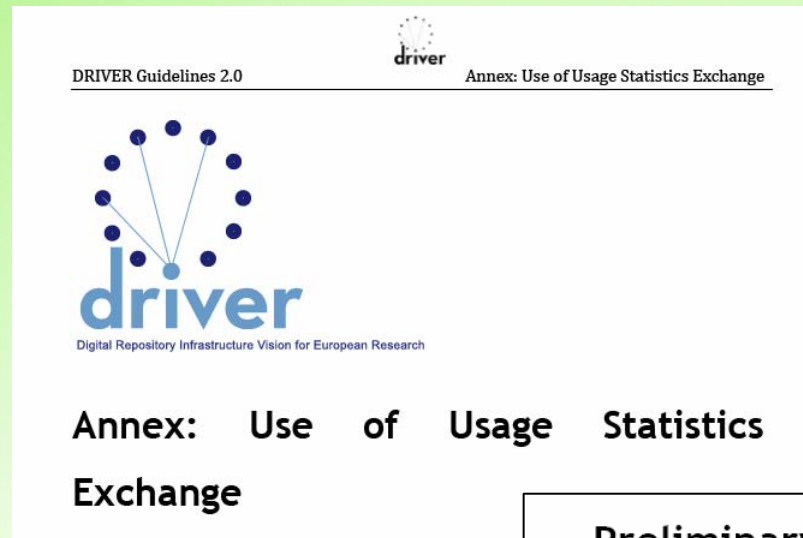
Los pasos que da la herramienta son:

- Lee las variables de configuración
- Crea un sólo fichero con todos los logs ordenados por fecha
- Crea un log de accesos sin entradas no válidas (sólo accesos con éxito por métodos GET o POST)
- Crea un log de descargas sin entradas no válidas (sólo descargas completas por métodos GET o POST)
- Elimina los robots obtenidos de las páginas <http://www.botsvsbrowsers.com/category/1/index.html>, <http://www.user-agents.org/allagents.xml> y del examen de los logs de accesos a las páginas del Consorcio Madroño
- Elimina dobles clicks de los accesos
- Elimina dobles clicks de las descargas
- Escribe las estadísticas por año mes y día

Tratamiento de los *logs* de salida del servidor Apache

[illegible]

Análisis de estadísticas de uso en las DRIVER Guidelines 2.0



Preliminary results of the project OA-Statistik

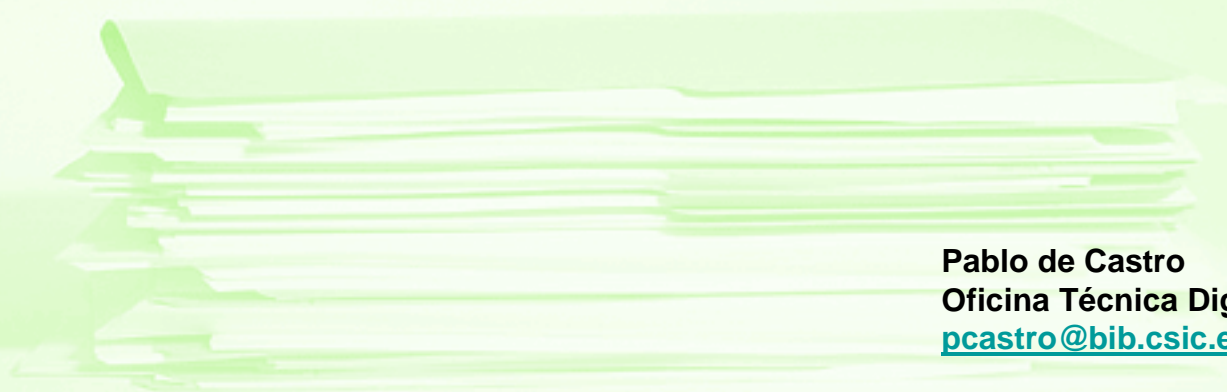
Goals of OA-Statistics

We aim to produce valid and reliable document usage statistics based solely on information gathered from the HTTP layer.

There are two main issues addressed by all existing standards which generate the bulk of the necessary corrections:

- Identification of non-human access
- Multi-Click correction

¡ Gracias !



Pablo de Castro
Oficina Técnica Digital.CSIC
pcastro@bib.csic.es